

# Abstracts

## Keynote 2

### Methodological approaches in occupational epidemiology

#### K2.1 PERSON-TIME AT RISK: UNIFYING CONCEPTS FROM AN OLD IDEA

D. Loomis. *Department of Epidemiology, University of Carolina at Chapel Hill, USA*

The analysis of person-time risk data from working populations dates from the earliest years of quantitative epidemiology: William Farr and other pioneering statisticians had begun to use census and death registration data to generate occupational mortality rates over 150 years ago. Initially person-time at risk was approximated by the number of people living in a given year (usually tabulated according to demographic categories defined by sex and age). Today we understand data of this type to estimate the rate of incidence, a fundamental epidemiological indicator of disease occurrence.

Although the incidence rate has natural and historical connections to the analysis of population health, modern epidemiology has tended to emphasise study designs—such as the clinical trial and the case control study—that treat time at risk in different, and less natural, ways. The parallel development of statistical methods has also favoured techniques adapted to these designs, notably the logistic and proportional hazards regression models. While these developments have undoubtedly advanced the state of the art, they have done so at a price. The challenges of articulating explanations of “odds” and “hazard” might be mentioned, but for occupational epidemiology the most significant concern is lack of clarity about the time related nature of exposure.

Occupational health researchers are usually interested in exposures whose level can change over time. When exposure changes in this manner, the risk of disease may also vary continuously in time as a function of current and previous exposure. However, common epidemiological designs and analytical techniques typically do not take this temporal variability explicitly into account. One side effect of this methodology is that the relevant exposure may be obscured, making it difficult for epidemiologists and exposure assessors to discuss, define, and collect the kind of data that are needed to quantify the time related association of disease with occupational agents.

In science, returning to old ideas sometimes helps to advance theory or technique, and in that way perhaps the old practice of person-time analysis can provide concepts to unify approaches to epidemiology and exposure assessment. Poisson regression has been used for some time to analyse incidence rate data from longitudinal cohort studies, but the conventional application of this method requires person-time and cases to be categorised, as in classical analyses of vital statistics. However, with the inexpensive computing power now available, Poisson regression models can be fit to individual units of person-time, stratified as finely as the analyst desires. The analysis of data in this form opens virtually infinite possibilities for considering exposure, its covariates, and disease occurrence in a time related manner. It also emphasises the value of collecting data that can characterise the exposure potentially relevant to risk for a particular unit of person-time. Detailed data on time varying exposures and appropriate analytical methods are both crucial to the analysis of disease latency, for example. Biologically based models can also be applied to refine the estimation of time related risk. Occupational studies oriented around these concepts have the potential to improve the assessment of risk and are strongly encouraged.

#### K2.2 BIAS FROM DEPENDENT MISCLASSIFICATION IN OCCUPATIONAL EPIDEMIOLOGY

P. Kristensen. *National Institute of Occupational Health, POB 8149 Dep, N-0033 Oslo, Norway*

The occupational epidemiologist has the fortune to study real problems in the real world. Alas, this is also our misfortune because we have to focus at all times on the internal validity of our results: can we trust our estimate

of association between exposure and outcome, or is it biased? Measurement error (misclassification for categorical variables) is one main source of biased results (information bias). The literature mainly describes two types: differential misclassification (the dirty type) and non-differential misclassification (the not-so-dirty type).

Differentiality presumes that only one of the two, exposure or outcome, is misclassified. Most likely we will have error in both variables, and this introduces a new problem: is the error (misclassification) dependent or not? Sometimes, we should ask this question. Unfortunately, epidemiologists are not in the habit of doing this although it is a well known issue in sociology and psychology. The definition is: *Errors in two variables are independent if the degree of error in one variable does not correlate with the degree of error in the other, if not, the errors are dependent.* An example: we want to examine the association between job demands and cardiac disease. We could measure both by objective means, but choose to ask participants to complete questionnaires, one on perceived demands at work and one on angina (Rose angina). Compared to the objective measures we are likely to encounter false positives and false negatives for both variables. It would not be implausible to assume that false positives for one variable would have an increased probability of being false positive on the other, and vice versa for false negatives. As can be understood, dependent misclassification will usually create falsely inflated estimates of association.

The combined error in the two variables must have a common source. Three common sources, often interacting, have been described. Firstly, personality trait: negative affectivity is a personality dimension being linked to dependent error. It is important to recognise that the whole spectre of this dimension contributes to the problem, not only those who are likely to overreport negative happenings. Secondly, transitory moods dependent on conditions in an investigation may influence results and introduce dependent error. Thirdly, investigation tools, in particular questionnaires, may have shortcomings that create dependent error.

Dependent error and corresponding bias are generated in the study design. Problems are usually but not always seen in cross sectional studies where the source of information on both exposure and outcome is study subjects, most commonly from questionnaires. Documented examples from occupational epidemiology are studies on psychosocial factors, on indoor air, and on musculoskeletal disorders.

What are the remedies to prevent bias from dependent error? First, we should recognise that not all surveys are suited for aetiological investigation. We might in fact experience that data generated from validated questionnaires, excellent from a descriptive point of view, produce flawed exposure-outcome associations. To avoid dependent error, we should break the bond between exposure and outcome information when designing studies. Improving our tools (questionnaires) could be another measure. In some situations, we could assess dependent error by estimating associations in population subsets, divided according to personality characteristics.

#### K2.3 CASE CONTROL STUDIES OF LOW EXPOSURES TO BENZENE: A DATA CONSISTENCY REVIEW

B. G. Miller<sup>1</sup>, W. Fransman<sup>2</sup>, H. Kromhout<sup>2</sup>, J. F. Hurley<sup>1</sup>, D. Heederik<sup>2</sup>, E. Fitzsimons<sup>3</sup>. <sup>1</sup>*Institute of Occupational Medicine, Edinburgh, UK;* <sup>2</sup>*Institute for Risk Assessment Sciences, University of Utrecht, the Netherlands;* <sup>3</sup>*Department of Haematology, University of Glasgow, UK*

**Introduction:** There have been four major case control studies of leukaemia in oil distribution and refinery workers, in the USA, Canada, Australia, and the UK. The results produced are discordant, in that the Australian study suggests evidence of leukaemia risks (excesses of exposure in cases over controls) at lower levels than the other studies. This implies differences in the observed exposure-relations between the studies.

**Methods:** This review set out to examine the existing studies, characterising their similarities and differences, and concentrating on those aspects that might contribute to differences such as those observed. It was important to consider the potential of many differences in methods and procedures for producing the observed differences. A second purpose was to assess the prospects for combining the data sets for a pooled analysis, which would have greater power than the individual studies. The review was carried out by a team of experienced epidemiologists and occupational hygienists, working closely with the principal investigators, and visiting each site to interview key workers

and to inspect research records. Visits, over the summer of 2004, followed a standardised protocol. The US study was evaluated from published results only, without a visit.

**Results:** The visits allowed documentation, description, and evaluation of the many common elements and differences between the studies. We found the procedures well documented and the data files in good order. Some missing data were identified in the UK study. While there were some differences in, for example, cohort inclusion, case ascertainment, etc, these did not seem likely explanations of the

observed differences between study results. There was a difference in the exposure distributions in the original cohorts, which may have been important for the power of the studies to detect exposure-response effects.

**Conclusion:** A pooled analysis may demonstrate that the results from the studies are not as discrepant as first thought. Its value would be enhanced by further work to demonstrate that the exposure assessments gave comparable results, although this would be hampered by the currently missing data.